ORIGINA L ARTICLE

Analysis of the positive predictive value of the subcategories of BI-RADS® 4 lesions: Preliminary results in 880 lesions


a Unidad de Radiología de la Mujer, Grupo Hospital de Madrid, Madrid, Spain
b Departamento de Estadística, Universidad CEU-San Pablo, Madrid, Spain
c Servicio de Radiodiagnóstico, Hospital Universitario Marqués de Valdecilla, Santander, Spain
d Unidad de Patología Mamaria, Hospital San Pedro, Logroño, Spain

Received 19 January 2011; accepted 11 April 2011

KEYWORDS
Breast tumors; Mammography; Breast ultrasonography; Positive predictive value; Relative risk

Abstract

Objectives: The positive predictive values (PPVs) of the subcategories of BI-RADS® 4 lesions (A/B/C) vary widely, and their correlation with specific descriptors has yet to be defined. We aimed to analyze the PPV of the subcategories and of the mammographic and ultrasonographic descriptors assigned to each.

Material and methods: We analyzed 880 histologically confirmed lesions prospectively classified as BI-RADS® 4 A/B/C between 2003 and 2010. The statistical analysis included significance tests, contingency tables, and relative risk (RR) ratios, calculated for 545 mammographic lesions and 627 ultrasonographic lesions.

Results: The PPV was 8.8% for subcategory 4A, 18.9% for subcategory 4B, and 58.3% for subcategory 4C. The correlation between PPV and lesions was what we expected, with three exceptions: (a) the PPV of 4A was greater than that of 4B in nodules that were irregular or had uncin cised margins on ultrasonography and in microcalcifications with segmental distribution on mammography, (b) BI-RADS® 3 lesions classified as BI-RADS® 4, and (c) identical lesions classified in distinct subcategories.

In the contingency table analysis, the mammographic lesions were 4B/C and the ultrasonographic lesions were 4B. On mammography, the RR was significant for nodules with irregular shape (RR = 3.205) and for those with spiculated margins (RR = 2.469), as well as for microcalcifications that were pleomorphic (RR = 2.531) or amorphous (RR = 0.334), and for those with segmental (RR = 1.895). On ultrasonography, the RR were significant for all the descriptors, with values greater than 1 for irregular shape (RR = 1.977) and uncin cised margins (RR = 2.277).

Conclusions: Our results corroborate previous reports. The exceptions can be explained by aspects related to variability and nonradiological factors that might influence the classification.
Análisis del valor predictivo positivo de las subcategorías BI-RADS®4: resultados preliminares en 880 lesiones

Resumen

Objetivos: Las subcategorías BI-RADS®4A/B/C comprenden amplios rangos de valores predictivos positivos (VPP) y no se ha definido su correlación con descriptores específicos. Nuestro objetivo es analizar el VPP de las subcategorías y los descriptor asignados a ellas en lesiones mamográficas y ecográficas.

Material y método: Analizamos 880 lesiones confirmadas histológicamente y subclasificadas prospectivamente como BI-RADS®4A/B/C entre 2003-2010. El estudio estadístico incluyó pruebas de significación, tablas de contingencia y estudio de riesgos relativos (RR) sobre 545 lesiones mamográficas y 627 ecográficas.

Resultados: Los VPP por subcategoría fueron 8,8%-4A, 18,9%-4B y 58,3%-4C. La correlación entre VPP y lesiones fue la esperada, excepto: VPP 4A>4B en nódulos ecográficos irregulares/márgenes no circunscritos y microcalcificaciones con distribución segmentaria, asignación de BI-RADS®4 a lesiones BI-RADS®3 y consideración de lesiones idénticas en distintas subcategorías.

En el estudio por tablas de contingencia, las lesiones mamográficas estuvieron en rangos de 4B/C y las ecográficas en 4B. Los RR fueron significativos en nódulos mamográficos para morfología irregular (RR = 3,205) y márgenes espículados (RR = 2,469), y para microcalcificaciones pleomórficas (RR = 2,531), amorfas (RR = 0,334) y distribución segmentaria (RR = 1,895).

En ecografía, los RR fueron significativos en todos los descriptor, con valores mayores de 1 en morfología irregular (RR = 1,977) y márgenes no circunscritos (RR = 2,277).

Conclusiones: Nuestros resultados concuerdan con los publicados. Las excepciones encontradas pueden justificarse por aspectos relacionados con la variabilidad y factores no radiológicos con posible influencia en la categorización y VPP. Es necesario elaborar modelos matemáticos que permitan la categorización objetiva e incluyan factores no relacionados con la imagen.

© 2011 SERAM. Publicado por Elsevier España, S.L. Todos los derechos reservados.

Introduction

The subdivision of BI-RADS® category 4 was included in the 4th edition of the Breast Imaging Reporting and Data System of the American College of Radiology (ACR)¹ in response to the need for risk stratification of a wide variety of lesions included between category 3 and 5, with positive predictive values (PPV) between 3 and 94%. BI-RADS® category 4 (suspicious abnormality-biopsy should be considered) is reserved for findings without the classic appearance of malignancy, but with higher concern of being a carcinoma than BI-RADS® category 3 findings. BI-RADS® subcategory 4A is assigned to findings requiring biopsy but having low suspicion of malignancy (PPV 3–10%). BI-RADS® subcategory 4B features lesions with intermediate suspicion of malignancy (PPV 11–50%) and requiring a close radiology–pathology correlation. BI-RADS® category 4C includes findings of moderate concern, but not classic for malignancy (PPV 51–94%). While histologic characterization is recommended for all categories, subcategorization helps radiologists, clinicians, and pathologists in the interpretation of results and in the decision-making based on radiology–pathology correlation after biopsy. Although this new subclassification has gained wide acceptance among radiologists, it has the same limitations and uncertainties attributed to the entire system since its introduction. The first limitation is variability, inherent to any description system based on perception. Its second limitation is the lack of correlation between subcategories and mammographic and ultrasonographic descriptors. The ACR recommends to assign category 4 by exclusion (PPV higher than BI-RADS® 3 PPV but lower than category 5), leaving the assignment to the different subcategories at the discretion of each center based on their experience.¹ Although some series have specifically addressed the analysis of the PPV based on the type of lesion and mammographic and ultrasonographic descriptors,²–⁴ no objective correlation of the factors that determine risk of carcinoma has been yet established. Thus, the assignment to a given subcategory remains a subjective matter, dependent on the experience of the radiologist and probably influenced by non-radiological factors. Additionally, validation studies, whether or not supported by mathematical models,⁵–⁸ require very large series of histologically confirmed lesions. In any case, variability is a limitation to the analysis of the results.

The aim of this study is to present our experience on the analysis of the PPV associated with BI-RADS® categories 4A/B/C assigned to mammographic and ultrasonographic descriptors.
Material and method

We conducted a multicenter study of 880 consecutive breast lesions (705 asymptomatic and 175 with associated clinical signs) classified as BI-RADS® categories 4A/4B/4C between 2003 and 2010. The regulations established by each center regarding data collection were followed and approval from the ethics committee was not considered necessary, as the study did not involve any experiments out of the standard practice for breast disease management.

All lesions were prospectively assigned to the subcategories, with subsequent histologic confirmation after core needle biopsy (12G in 383 patients, vacuum-assisted biopsy with 10-11G needle in 331, and both techniques in 166). Lesions confirmed as malignant received treatment (180 cases). Surgical biopsy was recommended in 73 cases for two reasons: high-risk histology results in 58 cases to rule out possible underestimation (the diagnosis of carcinoma was confirmed in 10), and disagreement between radiologic and pathologic findings, found in 15 cases (the diagnosis of carcinoma was confirmed in 3). Eleven of the 15 lesions with disagreement between radiologic and pathologic results were nodules, two were microcalcifications, one was a duct with solid content, and one was a focal asymmetry with associated microcalcifications. Three cases were assigned to subcategory 4A, four to 4B, and eight to 4C (the three confirmed carcinomas were in this subcategory). Eight benign lesions underwent surgery upon request of the patients. The results confirmed benignancy in all eight cases.

The 619 lesions that were not treated surgically and with benign outcome after biopsy were included in a follow-up protocol between 6 and 24 months. No carcinomas were confirmed.

Only 253 lesions (28.8%) were detected at mammography, 335 (38%) at ultrasound, and 292 (33.2%) using both techniques. Mammographic information was thus obtained from 545 and ultrasonographic from 627 lesions. All ultrasonographic images were assessed in real time.

The assignment of subcategories was exclusively based on the radiologist’s experience for two reasons: the BI-RADS system does not make explicit reference to the descriptors corresponding to each of the subcategories; the radiologists did not receive consensus training prior to data collection. Each case was interpreted by one single radiologist in a team of five, with full-time dedication to breast imaging (diagnosis and screening), and with 1–12-year experience in breast diagnosis. In total, 84.7% of cases were categorized by two readers with experience >10 years. Those cases identified at mammography and ultrasound were assessed by the same radiologist, who assigned one single subcategory per lesion on the basis of the most suspicious features. Upon categorization, the radiologists had information about the patient’s age and history, reason for consultation, and clinical characteristics in diagnostic cases. However, the possible influence of these aspects on categorization was not assessed at this stage of the study.

As each case was assessed by one single radiologist in a health care activity setting during 7 years, no specific study of the variability was carried out.

Methodology of analysis

The methodology involves two aspects:

1. Analysis of the PPV of the subcategories BI-RADS® 4 in the series.

   This is the study of the risk of carcinoma (overall and by prospectively assigned subcategory) for each type of lesion at mammography and ultrasound (microcalcifications, nodules, architectural distortion and focal asymmetry at mammography, and nodules at ultrasound) as well as for mammographic and ultrasound descriptors according to the 4th edition of the system. The descriptors corresponding to nodules and microcalcifications were applied to the analysis of the mammographic features. Nodules visualized on ultrasound were regarded as the descriptors for ultrasound morphology, border, and pattern. The rest of ultrasound descriptors were not analyzed because part of the series was imaged prior to the publication of the ultrasound lexicon and because no systematic collection was carried out after 2004. For this reason, the influence of these descriptors on the final category could not be analyzed. For ease of analysis, morphology descriptors have been classified as non-irregular and irregular, and borders, as circumscribed and uncircumscribed.

2. Retrospective study of the risk of carcinoma for individual descriptors and descriptors in combination in the series as a whole.

   At this stage, the PPVs were analyzed following histologic confirmation of the descriptors both separately and in combination (irrespective of the prospectively assigned subcategory), for microcalcifications at mammography and for nodules at mammography and ultrasound. Descriptors analyzed separately and in combination, with a PPV between 3% and 10% were assigned to BI-RADS® 4A; descriptors with PPV ranging 11–50% were classified into BI-RADS® 4B; and descriptors with PPV ranging 51–94% were classified into BI-RADS® 4C.

   The relative risk (RR) for each type of lesion was calculated on the mammograms, and the most suspicious descriptors and the most common descriptors were calculated using both mammography and ultrasound. The RR for each type of lesion was calculated taking the group of lesions that do not meet the analyzed condition as a reference (for instance, RR of carcinoma when the lesion is a nodule compared with the rest of lesions), and in descriptors, the RR of the descriptor analyzed compared with the rest of descriptors in the same group (for instance, RR of carcinoma for spiculated margins in nodules at mammography compared with the rest of margins).

Statistical study

Data were included in Excel tables, and were analyzed using SPSS software. Pearson’s chi-square test was used for statistical significance. The contingency coefficient was used to measure the degree of association between the variables of interest. To this aim, the risk was stratified from lower to higher level of suspicion of the descriptors analyzed. The order was established as follows:
Table 1  Lesions at mammography and ultrasound: number of lesions and positive predictive values (overall and by subcategory).

<table>
<thead>
<tr>
<th></th>
<th>4A</th>
<th>4B</th>
<th>4C</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lesions at mammography</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Microcalcifications</td>
<td>167 (10.2%)</td>
<td>79 (21.5%)</td>
<td>46 (56.5%)</td>
<td>292 (20.6%)</td>
</tr>
<tr>
<td>Nodule</td>
<td>50 (8%)</td>
<td>25 (20%)</td>
<td>35 (71.4%)</td>
<td>110 (30.9%)</td>
</tr>
<tr>
<td>Architectural distortion</td>
<td>15 (13.3%)</td>
<td>22 (13.6%)</td>
<td>35 (62.9%)</td>
<td>72 (37.5%)</td>
</tr>
<tr>
<td>Focal asymmetry</td>
<td>30 (10%)</td>
<td>26 (19.2%)</td>
<td>14 (57.1%)</td>
<td>70 (22.9%)</td>
</tr>
<tr>
<td>Adenopathy</td>
<td></td>
<td>1 (100%)</td>
<td>1 (100%)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>262 (9.9%)</td>
<td>152 (19.7%)</td>
<td>131 (62.6%)</td>
<td>545 (25.3%)</td>
</tr>
<tr>
<td><strong>Lesions at ultrasound</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodule</td>
<td>261 (8.8%)</td>
<td>126 (16.7%)</td>
<td>142 (59.2%)</td>
<td>529 (24.2%)</td>
</tr>
<tr>
<td>Others</td>
<td>54 (5.6%)</td>
<td>26 (23.1%)</td>
<td>18 (50%)</td>
<td>98 (18.1%)</td>
</tr>
<tr>
<td>Total</td>
<td>315 (8.3%)</td>
<td>152 (17.8%)</td>
<td>160 (58.1%)</td>
<td>627 (23.3%)</td>
</tr>
</tbody>
</table>

PPV for each type of lesion and subcategory between brackets.

- Nodules at ultrasound: morphology (non-irregular–irregular), margin (circumscribed–uncircumscribed), and ultrasonographic pattern (hyperechoic–heterogeneous–hypoechoic).

A confidence interval of 95% (CI 95%) was established to determine whether there are significant differences between PPVs.

The CI 95% was associated with the RR calculation and those features with values >1 were considered as increased risk of carcinoma compared with the rest of features in the same group (lesion types and descriptors).

Results

Analysis of the positive predictive value of the subcategories BI-RADS® 4 for carcinoma

Of the 880 lesions, 465, 228, and 187 were assigned to subcategory 4A (52.8%), 4B (25.9%), and 4C (21.3%), respectively.

A total of 193 carcinomas were histologically confirmed (180 by percutaneous biopsy and 13 by surgical biopsy), which means a PPV of 21.9% for the whole series (CI 95% [0.192; 0.248]). Arranged by subcategories, the PPV was 8.8% in 4A (41/465, CI 95% [0.064; 0.118]), 18.9% in 4B (43/228, CI 95% [0.140; 0.246]) and 58.3% in 4C (109/187, CI 95% [0.509; 0.655]).

Lesions at mammography

The most common findings were microcalcifications (53.6%, 529/545), followed by nodules (20.2%, 110/545), architectural distortion (13.2%, 72/545), and focal asymmetry (12.8%, 70/545). Table 1 summarizes the number of lesions assigned to each subcategory and the PPV arranged by type of lesion and subcategory. The statistical analysis demonstrated a significant association between the subcategory assigned and PPV (contingency coefficient 0.424 and p < 0.001).

Microcalcifications

Overall PPV was 20.6% (60/292, CI 95% [0.161; 0.257]) (Table 1). The statistical study revealed a significant association between morphology, the subcategory assigned (contingency coefficient 0.556 and p < 0.001), and risk of carcinoma (contingency coefficient 0.31 and p < 0.001). The relationship between the distribution with subcategory and risk of carcinoma was not statistically significant (contingency coefficient 0.236 and 0.146 with p = 0.11 and p = 0.12, respectively).

BI-RADS® subcategory 4A. A total of 167 lesions with a PPV of 10.2% (17/167, CI 95% [0.061; 0.158]) (Table 2) were classified in this subcategory. The most common were amorphous microcalcifications in a clustered distribution (85.6%, 143/167) with a PPV of 9.8% (14/143, CI 95% [0.055; 0.159]) (Fig. 1).

Figure 1  An example of amorphous microcalcifications with clustered distribution, classified as BI-RADS® subcategory 4A.
**Table 2**  Microcalcifications at mammography: distribution by subcategory and morphology and distribution descriptors.

<table>
<thead>
<tr>
<th>Morphology/distribution</th>
<th>Diffuse</th>
<th>Regional</th>
<th>Clustered</th>
<th>Linear</th>
<th>Segmented</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BI-RADS</strong> 4A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amorphous</td>
<td>0/1 (0)</td>
<td>1/4 (25)</td>
<td>14/143 (9.8)</td>
<td>0/3 (0)</td>
<td>1/8 (12.5)</td>
</tr>
<tr>
<td>Pleomorphic</td>
<td>0/1 (0)</td>
<td>1/6 (16.7)</td>
<td>0/1 (0)</td>
<td></td>
<td>0/1 (0)</td>
</tr>
<tr>
<td>Linear/branching</td>
<td>0/1 (0)</td>
<td>1/5 (20)</td>
<td>15/149 (10.1)</td>
<td>0/4 (0)</td>
<td>1/8 (12.5)</td>
</tr>
<tr>
<td>Punctuate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0/1 (0)</td>
</tr>
</tbody>
</table>

| **BI-RADS** 4B          |         |          |           |        |           |
| Amorphous                | 1/1 (100) | 0/1 (0) | 6/25 (24) | 1/8 (12.5) | 8/35 (22.9) |
| Pleomorphic              |         |          |           |        |           |
| Linear/branching         | 1/1 (100) | 0/1 (0) | 7/34 (20.6) | 0/5 (0) | 8/42 (19.1) |
| Punctuate                |         |          |           |        | 1/2 (50) |

| **BI-RADS** 4C          |         |          |           |        |           |
| Amorphous                | 0/1 (0) | 1/2 (50) | 1/1 (100) | 1/13 (7.7) | 17/79 (21.5) |
| Pleomorphic              | 0/1 (0) | 15/28 (53.6) | 5/8 (62.5) | 20/37 (54.1) |
| Linear/branching         | 0/1 (0) | 4/4 (100) | 0/1 (0) | 4/5 (80) |
| Punctuate                | 0/2 (0) | 16/31 (51.6) | 10/13 (77) | 26/46 (56.5) |

Carcinomas/cases. PPV per each combination of descriptors between brackets.

**BI-RADS** subcategory 4B. The total number of lesions in this subcategory was 79 with a PPV of 21.5% (17/79, CI 95% [0.130; 0.322]) (Table 2). Pleomorphic microcalcifications in a clustered distribution were the most common findings (43%, 34/79), with a PPV of 20.6% (7/34, CI 95% [0.087; 0.379]) (Fig. 2), followed by clustered amorphous microcalcifications (31.6%, 25/79), with a PPV of 24% (6/25, CI 95% [0.0936; 0.451]). Irrespective of the morphology, the PPV of the segmental distribution was 7.7% (1/13, CI 95% [0.002; 0.36]), which is lower than that obtained for the same descriptor in 4A (12.5%, 1/8, CI 95% [0.003; 0.526]).

**BI-RADS** subcategory 4C. A total of 46 lesions with a PPV of 56.5% (26/46, CI 95% [0.411; 0.710]) were classified in this category. Pleomorphic microcalcifications in clustered distribution were the most common (60.9%, 28/46) with a PPV of 53.6%, (15/28, CI 95% [0.339; 0.725]). The most suspicious descriptors analyzed separately were linear/branching morphology with a PPV of 80% (4/5) and segmental distribution (Fig. 3) with a PPV of 77% (10/13). In both cases, the CI 95% were very wide due to the reduced number of cases ([0.28; 0.99] and [0.46; 0.95], respectively), and thus, these descriptors were not considered representative.

**Nodules**

The nodules with and without microcalcifications were analyzed together. Overall PPV was 30.9% (34/110, CI 95% [0.224; 0.404]) (Table 1). Statistically significant differences were found in the descriptors of morphology and margin.

![Figure 2](image-url)  An example of pleomorphic microcalcifications with clustered distribution, classified as BI-RADS subcategory 4B.

![Figure 3](image-url)  An example of microcalcifications with pleomorphic and linear morphology and segmental distribution, classified as BI-RADS subcategory 4C.
for the subcategory assigned (contingency coefficients were 0.511 and 0.55, respectively, with $p < 0.001$) and risk of carcinoma (contingency coefficients were 0.383 and 0.31 with $p < 0.001$ and $p = 0.02$, respectively). The outcomes were not significant for density in either of the two aspects (contingency coefficients 0.191 and 0.032 with $p = 0.124$ and $p = 0.736$ for subcategory and risk of carcinoma, respectively).

**BI-RADS® subcategory 4A.** As many as 50 nodules with a PPV of 8% (4/50, CI 95% [0.0222; 0.192]) were included in this subcategory (Table 3). The most common combinations were round/oval/lobulated with circumscribed margins/occult (70%, 35/50) morphologies, with a PPV of 5.7% (2/35, CI 95% [0.007; 0.19]) (Fig. 4). Second in frequency were the same morphologies with ill-defined margins (30%, 15/50), with a PPV of 13.3% (2/15, CI 95% [0.017; 0.404]) (Fig. 5).

**BI-RADS® subcategory 4B.** A total of 25 nodules were included in this subcategory (Table 3) with a PPV of 20% (5/25, CI 95% [0.068; 0.407]). The most common combinations of morphologies were round/oval/lobulated with ill-defined margins (60%, 15/25), with a PPV of 13.3% (2/15, CI 95% [0.0165; 0.404]).

**BI-RADS® subcategory 4C.** As many as 35 nodules with a PPV of 71.4% (25/35, CI 95% [0.537; 0.853]) were classified into this subcategory (Table 3). The most common combinations featured were ill-defined/spiculated margins for all morphologies (82.8%, 29/35), with a PPV of 54.3% (19/35, CI 95% [0.367; 0.712]).

**Architectural distortion and focal asymmetry**

Overall PPV was 37.5% for architectural distortion (27/72, CI 95% [0.264; 0.497]) and 22.9% for focal asymmetry (16/70, CI 95% [0.137; 0.345]) (Table 1). By subcategories 4A/B/C, the PPV was 13.3% (2/15, CI 95% [0.016; 0.404]), 13.6% (3/22, CI 95% [0.0289; 0.349]), and 62.9% (22/35, CI 95% [0.450; 0.786]), respectively, for architectural distortion; and 10% (3/30, CI 95% [0.0211; 0.265]), 19.2% (5/26, CI 95% [0.065; 0.393]), and 57.1% (8/14, CI 95% [0.288; 0.823]), respectively, for focal asymmetry.

**Lesions at ultrasound (Table 1)**

The most common type of lesion was the nodule, which meant 84.4% of the total series (529/627). The remaining 15.6% (98/627) comprised a heterogeneous group of lesions (ductal disease, complex cysts, adenopathy, and other lesions), which were not considered representative because of their reduced number.

![Figure 4](http://example.com/nodule_mammography.png)

**Figure 4** Nodule at mammography with oval morphology and occult margins, classified as BI-RADS® subcategory 4A.
The nodules with and without microcalcifications were analyzed together. The PPV was 24.2% (128/529, CI 95% [0.206; 0.281]). The outcomes were statistically significant in terms of morphology and margins for subcategory (contingency coefficients 0.339 and 0.422, respectively, with p < 0.001) and risk of carcinoma (contingency coefficient 0.2 and 0.195, with p = 0.045 and p = 0.028, respectively). As for the echogenic pattern, a statistically significant correlation was found with the subcategory (contingency coefficient 0.223 with p = 0.017) and a non-significant correlation with the risk of carcinoma (contingency coefficient 0.075, with p = 0.52).

**Table 4** Nodules at ultrasound: distribution by subcategory and morphology and margin descriptors.

<table>
<thead>
<tr>
<th>Morphology/margin</th>
<th>Circumscribed</th>
<th>Uncircumscribed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BI-RADS® 4A</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-irregular</td>
<td>4/127 (3.2)</td>
<td>11/99 (11.1)</td>
</tr>
<tr>
<td>Irregular</td>
<td>0/5 (0)</td>
<td>8/30 (26.7)</td>
</tr>
<tr>
<td></td>
<td>4/132 (3)</td>
<td>19/129 (14.7)</td>
</tr>
<tr>
<td><strong>BI-RADS® 4B</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-irregular</td>
<td>6/20 (30)</td>
<td>9/53 (17)</td>
</tr>
<tr>
<td>Irregular</td>
<td>6/20 (30)</td>
<td>6/53 (11.3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15/106 (14.2)</td>
</tr>
<tr>
<td><strong>BI-RADS® 4C</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-irregular</td>
<td>11/15 (73.3)</td>
<td>28/48 (58.3)</td>
</tr>
<tr>
<td>Irregular</td>
<td>1/1 (100)</td>
<td>44/78 (56.4)</td>
</tr>
<tr>
<td></td>
<td>12/16 (75)</td>
<td>72/126 (57.1)</td>
</tr>
</tbody>
</table>

Carcinomas/cases. PPV per each combination of descriptors between brackets.
subcategories 4A and 4B, the irregular nodules with uncin-
circumscribed margins showed a lower PPV when assigned to
the subcategory 4B (11.3%, 6/53, CI 95% [0.0426; 0.230])
than when assigned to the 4A. Similarly, the PPV for uncin-
circumscribed margins—irrespective of their morphology—was
similar in both subcategories (Table 4).

**Bi-RADS® Subcategory 4C.** As many as 142 nodules (Table 4)
with a PPV of 59.2% (84/142, CI 95% [0.506; 0.674]) were
classified into this subcategory. The most common combi-
nation was uncircumscribed margin for all morphologies in
88.7% of the series (126/142) with a PPV of 57.1% (72/126,
CI 95% [0.480; 0.659]) (Fig. 7).

**Retrospective study of risk of carcinoma for
individual and clustered descriptors in the
whole series**

**Lesions on mammography**

The PPVs of the mammographic descriptors are included in
Table 5. Risk of carcinoma associated with descriptors ana-
lyzes separately and in combination was found in ranges
corresponding to subcategories 4B to 4C.

By type of lesions, the outcomes were statistically
significant in nodules (RR = 1.573, CI 95% [1.183; 2.091]
and microcalcifications (RR = 0.666, CI 95% [0.498; 0.892],
$p < 0.006$).

![Figure 7](image.png)

**Figure 7** Nodule at ultrasound with irregular morphology and
uncircumscribed margin, classified as Bi-RADS® 4C.

As for descriptors considered separately, significant out-
comes were reported for pleomorphic (RR = 2.531, CI 95%
[1.699; 3.769], $p < 0.001$) and amorphous (RR = 0.334,
CI 95% [0.222; 0.502], $p < 0.001$) morphology, as well as for
segmental distribution (RR = 1.895, CI 95% [1.163; 3.087],
$p < 0.017$) in microcalcifications. In nodules, irregular mor-
phology (RR = 3.205, CI 95% [2.018; 5.088], $p < 0.001$) and
spiculated margins (RR = 2.469, CI 95% [1.474; 4.135],

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Positive predictive value by combination of microcalcification and nodule descriptors at mammography.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microcalcifications</strong></td>
<td>Diffuse</td>
</tr>
<tr>
<td>Amorphous</td>
<td>100%</td>
</tr>
<tr>
<td>Pleomorphic</td>
<td>1/1</td>
</tr>
<tr>
<td>Linear/branching</td>
<td>0</td>
</tr>
<tr>
<td>Punctuate</td>
<td>0</td>
</tr>
<tr>
<td>Linear</td>
<td>50%</td>
</tr>
<tr>
<td>1/2</td>
<td>1/8</td>
</tr>
<tr>
<td><strong>Nodules</strong></td>
<td>Circumscribed</td>
</tr>
<tr>
<td>Round</td>
<td>33%</td>
</tr>
<tr>
<td>2/6</td>
<td>2/13</td>
</tr>
<tr>
<td>Oval</td>
<td>22%</td>
</tr>
<tr>
<td>2/9</td>
<td>9/24</td>
</tr>
<tr>
<td>Lobulated</td>
<td>20%</td>
</tr>
<tr>
<td>2/10</td>
<td>0/3</td>
</tr>
<tr>
<td>Irregular</td>
<td>24%</td>
</tr>
<tr>
<td>6/25</td>
<td>3/23</td>
</tr>
</tbody>
</table>

Carcinomas/cases.
Lesions at ultrasound

The PPVs for nodule descriptors were within the range of subcategory 4B (Table 6). RR was statistically significant (p < 0.001) for all descriptors considered separately, with values >1 for irregular morphology (RR = 1.977, CI 95% [1.462, 2.673]) and uncircumscribed margins (RR = 2.277, CI 95% [1.492, 3.476]), and values <1 for non-irregular morphology (RR = 0.506, CI 95% [0.374, 0.684]) and circumscribed margins (RR = 0.439, CI 95% [0.288; 0.67]).

Discussion

BI-RADS® category 4 encompasses a wide variety of lesions for which histologic characterization is recommended. From the first edition of the BI-RADS® system, a good number of series have shown a significant overlap of lesions assigned to each category, showing higher variability in intermediate categories (BI-RADS® 3 and 4) than in the most extreme categories (BI-RADS® 2 and 5). The definition of the specific characteristics of BI-RADS® 3 lesions proposed by Sickles and Varas was a significant step to establish objective criteria of suspicion, but their implementation has not become a standard, and their application in clinical practice is unclear yet. There was thus the need to subdivide the intermediate categories in order to reduce the wide range of suspicion. Mendez et al. suggested subdividing the BI-RADS® 3 category in 3A–3B. Recommendations would consist of short-term follow-up (nodules with PPV < 3%) and biopsy (microlcalcifications with PPV normally > 3% according to Mendez et al.) for 3A and 3B, respectively.

In response to this need for subdivision, the ACR classified BI-RADS® 4 into 4A/B/C in the 4th edition of the system, setting a PPV range between 3% and 94%. However, the PPV and features of each subcategory were not defined, and thus, the assignment criteria remain subjective and dependent on individual experience. As in previous editions, the system recommends initial histologic analysis, giving more weight to the subcategories in decision-making process according to the radiology–pathology correlation.

To our knowledge, our study presents the widest series, which specifically focuses on the analysis of the subcategories of BI-RADS® 4. The PPVs obtained by subcategory are statistically significant and within the ranges reported for mammographic and ultrasonographic lesions. However, in a closer analysis of the outcomes we found three remarkable aspects: the “paradoxical” results obtained between subcategories 4A and 4B for specific descriptors; consideration of mammographic and ultrasonographic lesions with BI-RADS® 3 features within the subcategories BI-RADS® 4; and assignment of different BI-RADS® subcategories 4 to the same lesions and combination of descriptors.

The “paradoxical” results were found between subcategories 4A and 4B in the segmental distribution of microlcalcifications and in irregular morphology and uncircumscribed margins, both separately and in combination, at ultrasound. In both cases, the PPV of subcategory 4A was higher than the same features subcategorized as 4B. Although the justification for these differences is likely to be multifactorial, variability studies have shown a lower inter- and intraobserver correlation in these descriptors. The Kappa index values reported for microlcalcification distribution range between moderate and poor (particularly in segmental distribution), with no significant improvement following formation. For the margins of nodules seen at mammography and ultrasound, the Kappa indexes are described as moderate-poor. This lack of agreement in medical literature is probably due to the difficulty in recognizing these descriptors and in correctly assigning the level of suspicion. In the series presented, each case was interpreted by only one radiologist. Since data were collected from consecutive cases, in a health care setting, and prospectively, the study was not designed to take into consideration the possible influence of the variability in outcomes.

The number of cases with probably benign typical features categorized as BI-RADS® 4 was low, but with PPV > 2%, it was increasing up to the expected percentages for each subcategory. Similarly, radiologically “identical” lesions were assigned to different subcategories with good correlation with the PPV. Conceptually, the BI-RADS® system bases the assignment of categories on morphologic criteria without considering so far the inclusion of non-radiological criteria. Nevertheless, a good number of series have looked into the possible influence of non-radiological factors, such as the presence of symptoms, changes over the follow-up interval, and history of risk factors. While it has not been shown to influence the risk of carcinoma in BI-RADS® 3 lesions, the presence of symptoms (essentially a palpable lesion) does seem to have a bearing on BI-RADS® 4. According to the outcomes reported by Kim et al. on a series of 519 lesions classified as BI-RADS® 4, the PPV in palpable lesions was 54%, in contrast with 16.8% in non-palpable lesions. This fact would warrant the assignment of a distinct subcategory based on the presence of symptomatology.

The outcomes are not that conclusive when the possible influence of the changes over the follow-up interval is analyzed. Accordingly, Lehman et al. found significant differences in PPV for BI-RADS® 3 lesions. In another study, 16.9% of the probably benign lesions that grew over the follow-up from a total of 4,514 cases from three series were malignant, with PPV within the ranges of subcategory 4B. This increase is more significant in microlcalcifications,
in which an increase in size during follow-up may yield PPV between 32% and 37%, compared with 0–12% in stable lesions. The difference between series may be related to heterogeneity in applying probably benign criteria, although most studies emphasize the importance of assessing developmental changes for risk estimation. Age, menopausal status, and family history have also been shown to have a significant impact on the risk of developing carcinoma and on risk estimation by the radiologist.

Another recently analyzed factor is the difference in PPV of nodules detected in locoregional staging according to the nodule location with respect to the carcinoma. Kim et al. demonstrate an increase in PPV of BI-RADS 3 and BI-RADS 4 lesions between 4.2 and 21.2% and between 22.6 and 61.1%, respectively, depending on whether the lesion is located in the contralateral breast or in the same quadrant of the carcinoma.

Although there are no specific studies on the influence of non-radiological factors, our outcomes might well reflect the subjective component that these factors may have on categorization by the radiologist—who had clinical information at the time of the study—when assigning similar lesions to different subcategories, with good correlation and expected PPV.

Contingency tables are of help in category assignment because they enable to prospectively calculate the PPV of descriptors separately and in combination according to the outcomes of wide series of confirmed cases. Since the series reported by Liberman et al. and Orel et al., a good number of studies have approached the analysis of the risk of carcinoma by relying on contingency tables, yielding similar results with regard to identification of features with higher PPV or NPV (negative predictive values).

In our series, most descriptors, both separate and in combination, of nodules and microcalcifications at mammography and all the nodules at ultrasound are within the PPV range assigned to subcategory 4B. Although in a small number, lesions with typical features of high suspicion (irregular morphology with spiculated margin in nodules and linear/branching morphology in microcalcifications) have been categorized as BI-RADS 4 lesions. However, PPV was lower than 95% in all cases.

The RR study has been of use in identifying the lesions and descriptors regarded as risk factors for malignancy (RR values and CI 95% > 1 with statistical significance) or as low risk factors in comparison with the rest of lesions and/or descriptors in their same group (RR and CI 95% < 1 with statistical significance). By type of lesion at mammography, the nodules show RR and CI 95% values > 1, which is indicative of a significantly higher risk of carcinoma when the lesion classified as BI-RADS 4 is a nodule than when the lesion corresponds to microcalcification with RR and CI 95% < 1. With respect to the descriptors considered separately, the RR and CI 95% are > 1 with statistical significance for pleomorphic morphology and segmental distribution in microcalcifications, and for irregular morphology and spiculated margins in nodules, which are therefore regarded as increased risk factors in those lesions exhibiting these features. However, amorphous microcalcifications, the most commonly classified as BI-RADS 4, have RR and CI 95% values < 1. Therefore, and as long as they are not in association with segmental distribution, the identification of amorphous microcalcifications would be indicative of a lower level of suspicion than other morphologies of microcalcifications.

At ultrasound, clustering morphology and margin descriptors of nodules have been of use to obtain significant outcomes in all individual descriptors. Consequently, irregular morphology and circumscribed margins can be considered as increased risk factors, and non-irregular morphology and circumscribed margins, as low risk factors.

Probably, the most useful advantage provided by contingency tables concerns the design of mathematical models of risk prediction. A variety of models have been described in artificial neuronal networks (ANN), Bayesian networks, and risk ratios. All these models require extensive databases, as well as the possibility of growth and “learning”. Most models build upon an exclusively radiological dataset. Nevertheless, the addition of clinical and epidemiological factors as well as the radiologist’s opinion are showing a better correlation with suspicion level, leading to a preference for the design of mixed models. The outcomes yielded by this series are the building blocks for the design of a mixed model that would encompass the whole range of BI-RADS categories 3–5 on mammography and ultrasound.

Our study has four limitations. The first one is the prospective design, which is not oriented to agreement assessment, and thus, determination of the degree of impact of this assessment on the outcomes is not possible. Variability is an unavoidable phenomenon related to perception and to the lack of accuracy of the lexicon used. Accordingly, terms such as “amorphous” and “pleomorphic” are hard to be narrowed down, and justify the lack of agreement between observers. The second limitation is the number of cases, which prevented us from obtaining significant outcomes for certain features. Availability of wide series is the most serious limitation in the series reported, showing greater difficulty with a higher number of possible combinations. This fact translates in our study as wide CI 95% in PPV with a reduced number of cases, which renders some outcomes unrepresentative. The third limitation is the combined evaluation of mammographic and ultrasonographic features in the assignment of a single level of suspicion in lesions detected by both techniques. This fact, however, has had an impact on the PPV by subcategory. The fourth and final limitation concerns the lack of systematic collection of all of the ultrasonographic descriptors and the disregard of factors unrelated to imaging. Both factors have undoubtedly limited the possibilities of our study.

In conclusion, the subdivision of BI-RADS category 4 is useful, and the outcomes are reproducible both in terms of PPV and of features of lesions. However, there is a major subjective component, most likely to be associated with factors unrelated to imaging. These factors have an impact on categorization, and determine the PPV in each subcategory. However, they have as yet not been considered by the ACR, nor are they systematically accounted for in clinical practice. Further studies are thus required on wider series including categories 3 and 5. These studies should be built upon mathematical models considering both radiological features and factors unrelated to imaging. Such models should help in the decision-making for the indication
for biopsy and in the protocol of action after histologic outcomes.

Authorship

1. Responsible for the integrity of the study: MTT.
2. Conception of the study: MTT.
3. Design of the study: All the authors.
4. Acquisition of data: All the authors.
5. Analysis and interpretation of data: MTT, JMCR, MSG, and MSM.
6. Statistical analysis: JMCR.
7. Bibliographic search: MTT.
8. Writing of the manuscript: MTT.
9. Critical review with intellectually relevant contributions: JMCR, MSG, and MSM.
10. Approval of the final version: All the authors.

Conflict of interest

The authors declare not having any conflict of interest.

References

Analysis of the positive predictive value of the subcategories of BI-RADS® 4 lesions in 880 lesions

